MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2014) **14**, 508–518 doi: 10.1111/1755-0998.12208

Recovery of DNA barcodes from blackfly museum specimens (Diptera: Simuliidae) using primer sets that target a variety of sequence lengths

L. M. HERNÁNDEZ-TRIANA,* S. W. PROSSER,* M. A. RODRÍGUEZ-PEREZ,† L. G. CHAVERRI,‡ P. D. N. HEBERT* and T. RYAN GREGORY*

**Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario N1G 2W1, Canada, †Centro de Biotecnología Genómica, Instituto Politécnico Nacional. Blvd. del Maestro esq. Elias. Piña, Cd. Reynosa, Tamaulipas, Mexico, ‡Instituto Nacional de Biodiversidad, Avenida Santo Domingo de Heredia, Aptdo Postal 23-3100, San José, Costa Rica*

Abstract

In this study, we evaluated the efficacy of various primers for the purpose of DNA barcoding old, pinned museum specimens of blackflies (Diptera: Simuliidae). We analysed 271 pinned specimens representing two genera and at least 36 species. Due to the age of our material, we targeted overlapping DNA fragments ranging in size from 94 to 407 bp. We were able to recover valid sequences from 215 specimens, of which 18% had 500- to 658-bp barcodes, 36% had 201- to 499-bp barcodes and 46% had 65- to 200-bp barcodes. Our study demonstrates the importance of choosing suitable primers when dealing with older specimens and shows that even very short sequences can be diagnostically informative provided that an appropriate gene region is used. Our study also highlights the lack of knowledge surrounding blackfly taxonomy, and we briefly discuss the need for further phylogenetic studies in this socioeconomically important family of insects.

Keywords: blackflies, COI, DNA barcoding, museum specimens, Simuliidae

Received 24 August 2012; revision received 20 November 2013; accepted 22 November 2013

Introduction

Blackflies (Diptera: Simuliidae) comprise 26 genera and an estimated 2154 valid species (Adler & Crosskey 2013). In the majority of blackfly species, the female requires a bloodmeal for egg maturation, and it is this requirement that makes blackflies commercially and medically important. The most important human parasites transmitted by blackflies are the nematodes *Onchocerca volvulus* (Leuckart), the causative agent of river blindness, and *Mansonella ozzardii* Manson, which causes Mansonelliasis or 'serous cavity filariasis' (Shelley *et al*. 2010). In Latin America, blackflies are thought to be responsible for outbreaks of endemic pemphigus foliaceus (Eaton *et al*. 1998) and to be the aetiological agent of the Altamira haemorrhagic syndrome (Pinheiro *et al*. 1974, 1986; Pinheiro 1983). Blackflies also transmit pathogens to domestic livestock, resulting in increased mortality, reduced weight, decreased milk production and malnutrition (Adler *et al*. 2004; Currie & Adler 2008).

Correspondence: Luis M. Hernández-Triana and Sean W. Prosser, Fax:+1 519 824 5703; E-mails: lhernand@uoguelph.ca, sprosser@uoguelph.ca

In addition to the medical and commercial importance of some species, blackflies are an environmentally important group because of their role as 'keystone' organisms in the ecology of freshwater ecosystems. Specifically, blackfly larvae have the ability to filter dissolved organic matter, making it available to the downstream food chain (Malmqvist *et al*. 2001, 2004; Currie & Adler 2008). Blackfly larvae are also an important food source for salmonid fish and benthic invertebrates such as stoneflies (Currie & Adler 2008). Additionally, the susceptibility of immature blackfly larvae to environmental pollutants (both organic and inorganic) makes them very useful as bioindicators of landscape degradation (Feld *et al*. 2002; Lautenschalger & Kiel 2005; Pramual & Kuvangkadilok 2009; A. J. Shelley, personal communication).

Because of their significant medical, commercial and environmental importance, blackflies are one of the target groups involved in developing the Barcode of Life Database (BOLD- www.barcodinglife.com), a DNA barcode reference library (Hebert *et al*. 2003a,b, 2004; Ratnasingham & Hebert 2007). In order to develop a DNA barcode library for blackflies, it is very desirable to have freshly collected specimens from which one can easily obtain a full-length barcode fragment (658 bp) of the

mitochondrial cytochrome *c* oxidase subunit I (COI) gene (Folmer *et al.* 1994; Hebert *et al.* 2003a,b). However, there are various reasons why it is also desirable to obtain DNA barcodes from specimens preserved in museum collections. For example, this allows for the creation of linkages between contemporary and historical collections, which is useful for studying changes in species distributions over time (e.g. as a result of climate change) (Bucklin *et al.* 2010) or in assessing patterns of change in genetic diversity within populations (e.g. Watts *et al.* 2007). Barcoding museum specimens is also useful for resolving taxonomic uncertainty regarding correspondence to type specimens when cryptic species diversity is uncovered, for linking life-history stages within species and for integrating molecular and morphological data (e.g. Mandriolli *et al.* 2006; Hausmann *et al.* 2009; Rivera & Currie 2009). In some cases, museum specimens provide the only available samples for rare or otherwise difficult to acquire species (Schander & Halanychi 2003).

Unfortunately, the use of museum specimens to generate DNA barcodes can be challenging due to factors such as DNA degradation, contamination and uncertainty regarding details of specimen collection and preservation. It is well documented that collecting and preservation techniques can have a direct effect on the quality of DNA yields from museum specimens (Koch *et al.* 1998; Quicke *et al.* 1999; Dean & Ballard 2001; Knölke *et al.* 2005; Rivera & Currie 2009; Shokralla *et al.* 2011). A recent large-scale study by Hebert *et al.* (2013) tested the ability of various universal insect primers to recover barcode sequences from ~40 000 museum preserved specimens of Lepidoptera, ranging in age from approximately 10–90 years. In agreement with our personal observations, Hebert *et al.* found that recovery of full-length barcodes (i.e. 658 bp) via a single primer set is not feasible for specimens older than ~15 years. The analysis of preserved blackfly material is further complicated by the presence of parasites in the gut and pigments in the head which are known to inhibit PCR (Koch *et al.* 1998).

To overcome some of the problems highlighted above, specific molecular approaches can be employed to recover a short amplicon of the COI region (Hajibabaei *et al.* 2006; Meusnier *et al.* 2008; Baird *et al.* 2011; Hajibabaei & McKenna 2012). For example, the use of 'mini barcodes' in concert with next-generation DNA sequencing approaches (Shokralla *et al.* 2011), studies of lesions in DNA extracts from preserved samples (Zimmerman *et al.* 2008) and studies from specimens preserved in Carnoy's solution (Conflitti *et al.* 2010; Pramual *et al.* 2011) are currently being investigated. Unfortunately, our analysis of blackfly COI sequences (see 'Primer analysis and selection') suggested that standard 'mini-barcode' primers will not function efficiently on members of Simuliidae. In the present study, we tested the ability of several primer sets, targeting various fragment lengths, to recover barcode sequence from museum specimens of Simuliidae. The results of our study provide useful guidelines for the application of DNA barcoding when working with fresh and/or preserved blackflies and highlight the importance of primer selection and design when working with archived specimens.

Materials and methods

Source and preparation of museum specimens

In this study, we analysed 271 pinned specimens representing two genera (*Gigantodax* and *Simulium*) and approximately 36 species (Table 1). All blackfly specimens analysed in this study were from the collections of the Natural History Museum in London, England. Most of the specimens were >10 years old (mean = 26 years; median = 28 years, mode = 28 years). In order to avoid some of the problems outlined in Hernández (2011) with regard to using archived blackfly specimens for molecular analysis, specimens were only used when it was clear that they had been collected in accordance with the protocols outlined by R.W. Crosskey and A.J. Shelley (see Hernández 2007, 2011). For consistency, priority was given where possible to specimens that had been identified by L. M. Hernández-Triana and other Simuliidae taxonomic specialists such as R.W. Crosskey, A.J. Shelley, B. Cheke and R.J. Post. All specimens of *S. damnosum* were identified by R. W. Crosskey, B. Cheke and R.J. Post and re-examined by L. M. Hernández-Triana; the remainder species/specimens were identified by A. J. Shelley and L. M. Hernández-Triana and re-examined by the first author in 2012 (see Table 1). Emphasis was placed on sampling pinned, link-reared adults, but biting females of a known vector species or other anthropophilic species (biting pests) were also sampled. In addition, we sampled specimens of species that were poorly represented within BOLD such as *S. wolffhuegeli* and species belonging to the genus *Gigantodax*, which are very difficult to collect in the wild (Table 1).

Two to three legs were removed from each specimen for DNA extraction, and the remainder of the body was retained as a voucher. The leg samples were deposited into 96-well plates, and DNA extraction, PCR and sequencing were performed at the high-throughput facility at the Biodiversity Institute of Ontario (BIO), University of Guelph, Canada, following established protocols outlined by Ivanova *et al.* (2006) (see below for details). As an additional e-voucher, a digital image of each specimen was taken using either a standard digital camera or a Leica compound microscope equipped with a Z-stepper. Detailed specimen data

Table 1 List of genera, species, number of specimens, country of provenance and collection year of Simuliidae species

Genera	Species (No. of specimens)	Collection Year (No. of specimens)	Country (No. of specimens)
Gigantodax	*abalosi* (3)	1984 (3)	Ecuador (3)
Gigantodax	*basinflatus* (1)	1984 (1)	Ecuador (1)
Gigantodax	*cervicornis* (3)	1985 (3)	Ecuador (3)
Gigantodax	*flabellus* (5)	1984 (3); 1985 (2)	Ecuador (5)
Gigantodax	*lazoi* (6)	1984 (6)	Ecuador (6)
Gigantodax	*cormonsi* (1)	1979 (1)	Peru (1)
Gigantodax	*zumbahuae* (3)	1984 (3)	Ecuador (3)
Simulium	*nr. albilineatum* (1)	1979 (1)	Peru (1)
Simulium	*callidum* (7)	2001 (7)	Belize (7)
Simulium	*chaquense* (5)	1980 (5)	Argentina (5)
Simulium	*damnosum s.l.* (16)	1980 (12); 1981 (4)	Ghana (5); Benin (4); Togo(7)
Simulium	*dinelli* (5)	1971 (5)	Peru (5)
Simulium	*escomeli* (6)	1983 (6)	Ecuador (6)
Simulium	*exiguum s.l.* (13)	1984 (5); 1985 (3); 1998 (5)	Ecuador (8); Argentina (5)
Simulium	*ganalesense* (8)	2001 (8)	Belize (8)
Simulium	*gonzalezi* (13)	1981 (8); 2001 (5)	Belize (5); Ecuador(8)
Simulium	*haematopotum* (13)	1985 (3); 1990 (6); 2001 (4)	Belize (4); Dominican Republic (6); Panama (3)
Simulium	*ignescens* (7)	1984 (7)	Ecuador (7)
Simulium	*inaequale* (1)	1988 (1)	Ecuador (1)
Simulium	*lutzianum s.l.* (24)	1983 (5); 1984 (7); 1985 (9); 1988 (2); 1989 (1)	Ecuador (24)
Simulium	*nr. lutzianum s.l.* (7)	1984 (7)	Ecuador (7)
Simulium	*metallicum s.l.* (26)	1961 (4); 1978 (1); 1979 (2); 1983 (1); 1985 (8); 1986 (4); 1993 (2); 2001 (4)	Belize (5); Colombia (8); Ecuador (7); Panama (2); Venezuela (4)
Simulium	*nr. metallicum s.l.* (1)	1983 (1)	Ecuador (1)
Simulium	*ochraceum s.l.* (34)	1981 (6); 1984 (2); 1985 (4); 1987 (3); 1992 (9); 1996 (4); 2004 (6)	Colombia (2); Dominican Republic (9); Ecuador (14); Mexico (5); Puerto Rico (4)
Simulium	*perflavum* (5)	1970 (5)	Guyana(5)
Simulium	*nr. paynei s.l.* (1)	Unknown	Ecuador (1)
Simulium	*quadrivittatum* (6)	1990 (1); 1992 (2); 2001 (3)	Belize (3); Dominican Republic (3)
Simulium	*roraimense* (10)	1984 (10)	Ecuador (10)
Simulium	*samboni (10)*	1985 (5); 2001 (5)	Belize (5); Panama (5)
Simulium	*nr. spinifer* (6)	1984 (6)	Ecuador (6)
Simulium	*tarsatum* (5)	1937 (1); 1984 (4)	Ecuador (4); Trinidad and Tobago (1)
Simulium	*nr. tarsatum* (7)	1937 (3); 1985 (4)	Colombia (4); Trinidad and Tobago (3)
Simulium	*wolffhuegeli* (9)	1998 (9)	Argentina (9)
Simulium	*nr. wygodzinskyorum* (3)	1983 (3)	Ecuador (3)

records, sequence information (including trace files) and digital images were uploaded to BOLD under the project 'Old samples of blackflies (Simuliidae) held at the NHM Collection_2011, 2012 [Project Code: NHMBF]'. The DOI link for the data set for this project is dx.doi.org/10.5883/DS-NHMBF13. All sequences were also submitted to GenBank (Accession nos. KF839950 – KF840065). A list of genera, species, number of specimens, country of provenance and year of collection is given in Table 1.

Primer analysis and selection

To recover DNA sequences from archived specimens, previous studies (e.g. Hajibabaei *et al.* 2006; Meusnier *et al.* 2008) have utilized primer sets that amplify small overlapping fragments, which can be combined during *post hoc* analysis to form a full-length 658-bp COI barcode. For example, a recent study involving the barcoding of ~40 000 museum-housed Lepidoptera specimens utilized universal primer sets that amplify a 307-bp and a 407-bp fragment, which collectively span the entire 658-bp COI barcode region (Hebert *et al.* 2013; Table 2). Preliminary analyses at BIO have shown that the same primer sets can be used with many dipteran specimens; however, analysis of 1647 blackfly sequences downloaded from the BOLD database suggested that these primers may not bind efficiently to all blackfly DNA templates due to simuliid-specific polymorphisms within the primer-binding sites (not shown). We

Table 2 Primers used in this study. M13 tails (shown in bold) were added to primers, but only in cases where they were found to offer an advantage during sequencing

Cocktail	Primer	Sequence (5′ 3′)	References
C_LepFolF	LepF1	ATTCAACCAATCATAAAGATATTGG	Hebert *et al*. 2003a,b;
	LCO1490	GGTCAACAAATCATAAAGATATTGG	Folmer *et al*. 1994;
C_LepFolR	LepR1	TAAACTTCTGGATGTCCAAAAAATCA	Hebert *et al*. 2003a,b;
	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	Folmer *et al*. 1994;
n/a	MLepF1	GCTTTCCCACGAATAAATAATA	Hajibabaei *et al*. 2006;
n/a	MLepF2_t1	**TGTAAAACGACGGCCAGT**GCWTTCCCM CGWATAAATAATATAAG	Hebert *et al*. 2013;
n/a	MLepR2	GTTCAWCCWGTWCCWGCYCCATTTTC	Hebert *et al*. 2013;
n/a	MLepR3	GCTAARTGWARDGAAAAAATWGC	This study
C_microLepF1_t1	microLepF2_t1	**TGTAAAACGACGGCCAGT**CATGCWTTTAT TATAATTTTYTTTATAG	Hebert *et al*. 2013; Messing 1993
	microLepF3_t1	**TGTAAAACGACGGCCAGT**CATGCWTTT GTAATAATTTTYTTTATAG	Hebert *et al*. 2013; Messing 1993
C_TypeF1	TypeF1	ATTAGGAGCWCCWGATATRGC	Hebert *et al*. 2013;
	TypeF2	ATTAGGAGCWCCWGATATAGC	Hebert *et al*. 2013;
C_TypeR1	TypeR1	GGAGGRTAAACWGTTCAWCC	Hebert *et al*. 2013;
	TypeR2	GGAGGGTAAACTGTTCAWCC	Hebert *et al*. 2013;
	TypeR3	GGTGGATAAACAGTTCAWCC	Hebert *et al*. 2013

Primers used in cocktails were mixed in equal ratios; n/a – Not applicable.

Table 3 Primer sets used to obtain short barcode sequences from Simuliidae museum specimens

Primer set	Fragment size	PCR success rate	Sequencing success rate
C_LepFolF + MLepR2	307 bp	11.8%	11.4%
C_LepFolF + MLepR3	388 bp	25.8%	19.2%
MLepF1 + C_LepFolR	407 bp	44.3%	40.6%
MLepF2_t1 + C_LepFolR	403 bp	63.8%	43.5%
C_microLepF1_t1 + C_TypeR1	164 bp	n/a	80.8%
C_TypeF1 + C_TypeR1	94 bp	n/a	49.1%

The size of the corresponding amplicon and overall PCR and sequence recovery success rates are shown; n/a, not applicable.

therefore opted to test both the 307- and 407-bp primers alongside two counterpart primer sets that were predicted to anneal more efficiently to simuliid DNA. These primers amplify a 388- and 403-bp fragment (Table 3), and like the 307- and 407-bp fragments, overlap to form the full-length 658-bp COI barcode region. Specifically, the 388-bp primers were predicted to anneal significantly more efficiently to blackfly DNA compared with their counterpart 307-bp primers, whereas the 403-bp primers were predicted to anneal only slightly more efficiently to blackfly DNA compared with their counterpart 407-bp primers. Due to the possibility that the DNA from archived blackfly specimens may be too degraded to recover 300- to 400-bp fragments, we also processed our specimens with two additional primer sets that amplify a 164- and 94-bp fragment of the COI barcode region (Table 3). Overall, each specimen was processed with six primer sets (Table 2, Table 3, Fig. 1).

DNA extraction, PCR amplification and sequencing

DNA extraction and purification was performed using standard BIO protocols (Ivanova *et al*. 2006). For PCR amplification, 2 μL of purified DNA was added to a 10.5 μL PCR reaction consisting of 5% D-(+)-trehalose dihydrate (Fluka Analytical), 1X KAPATaq buffer (KAPA Biosystems), 2.5 μM $MgCl_2$ (Invitrogen), 0.1 μM of each primer, 0.05 mM dNTP (KAPA Biosystems), and 0.3 units of KAPATaq DNA Polymerase (KAPA Biosystems) for a total reaction volume of 12.5 μL. To amplify the 307-, 388-, 403- and 407-bp fragments, PCRs were thermocycled at 94 °C for 1 min, five cycles of 94 °C for 40 s, 45°C for 40 s, 72 °C for 1 min, followed by 35 cycles of 94 °C for 40 s, 51 °C for 40 s, 72 °C for 1 min and a final extension of 72 °C for 5 min. To amplify the shorter 164- and 94-bp fragments, PCRs were thermocycled for 40 cycles of 94 °C for 40 s, 51 °C for 40 s, 72 °C for 30 s and a final extension of 72 °C for 5 min. PCR products were visualized on a 2% agarose E-gel® 96 precast gel (Invitrogen). Due to their small size and susceptibility to contamination, the 164- and 94-bp PCR products were not visualized on an agarose gel, but directly sequenced. Samples were cycled using standard CCDB sequencing reactions (Ivanova *et al*. 2006) and bidirectionally

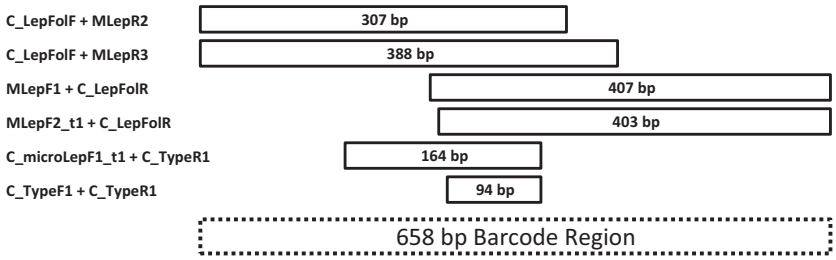

Fig. 1 Relative location, within the COI barcode region, of PCR fragments generated by each primer set.

sequenced on an ABI 3730XL capillary sequencer (Applied Biosystems). The resulting traces were edited using CODONCODE v. 3.0.1 (CodonCode Corporation, Dedham, Massachusetts).

Sequence analysis

Sequence analysis was performed using the analytical functions on BOLD. Recovered sequences were uploaded to BOLD and aligned using the built-in MUSCLE (Edgar 2004) algorithm, selected to include sample ID numbers and sequence length in labels and limited to sequences greater than 50 bp. A neighbour-joining (NJ) analysis was performed (Saitou & Nei 1987) using the K2P distance model (Kumar *et al.* 2007, 2012; Tamura *et al.* 2007, 2011) to validate the sequences by comparing the clustering patterns based on the barcode sequences to those expected based on morphological identifications. Inferred support for the NJ clustering patterns was obtained by performing a bootstrapped analysis in MEGA as detailed in Hernández-Triana *et al.* (2012).

Results

Table 3 shows the overall PCR and sequencing success rates for each primer set. The 388-bp fragment had an average success rate 7.8% higher than that of the 307-bp fragment, whereas the 403-bp fragment had an average success rate 2.9% higher than that of the 407-bp fragment (Table 3). These results correspond with our predictions that the 307- and 407-bp primers will not anneal to blackfly COI sequences as well as the 388- and 403-bp primers. After combining all overlapping fragments from each specimen, 12.5% of the specimens had full-length (658 bp) barcodes and approximately 30.3% had between 300 and 657 bp. Overall, 57% of the specimens did not provide sequence data.

To gain some sequence data from the samples that failed, we performed PCR on all 271 specimens using primers that amplify 164- and 94-bp fragments (Fig. 1). Table 3 shows the overall success rate of each of these primer sets. The 164-bp primers were able to recover sequences from 80.8% of the specimens, whereas the 94-bp primers recovered sequences from only 49.1% of the specimens. Of the sequences recovered by the 164-bp primers, approximately 96.9% were 164 bp. Conversely, of the sequences recovered by the 94-bp primers, only 41.8% were 94 bp, with the remaining 58.2% falling between 93 and 65 bp.

After combining the sequence data from all fragments and validating the resulting barcodes, we were able to recover high-quality sequences from 215 (80%) of our specimens. Of these, approximately 18% had full-length, or near full-length, barcodes (500–658 bp), 36% had barcodes ranging from 499 to 201 bp and 46% had barcodes less than or equal to 200 bp. The majority of our sequences were recovered from specimens collected between 10 and 35 years ago, although we did recover barcode sequences from specimens up to 51 years old.

Neighbour-joining analysis of our data set showed that specimens from the same species cluster together regardless of sequence length (Fig. 2). This observation lends credence to the notion that even short barcodes can be diagnostically relevant, provided that the gene region is chosen carefully (Hajibabaei *et al.* 2006; Meusnier *et al.* 2008). With the exception of a single grouping of *Simulium ochraceum s.l.* from Mexico, the majority of the specimens clustered as expected based on their morphological identification with strong bootstrapped support in most cases (Fig. 2). The five species complexes included in this study (*S. damnosum s.l.*, *S. exiguum s.l.*, *S. lutzianum s.l.*, *S. metallicum s.l.* and *S. ochraceum s.l.*) showed deep splits in the NJ tree, indicating high levels of genetic variation within each complex. Three species (*Gigantodax flabellus*, *G. lazoi* and *S. ignescens*) had levels of intraspecific divergence suggestive of cryptic diversity (e.g. Rivera & Currie 2009; Hernández 2011; Hernández-Triana *et al.* 2012), but further study is necessary to confirm this hypothesis.

Discussion

In studies that make use of old/degraded DNA, success rates are generally inversely proportional to the size of the target amplicon and highly dependent on the use of effective primers (Dean & Ballard 2001; Van Houdt *et al.* 2010; Bluemel *et al.* 2011). In this study, we utilized

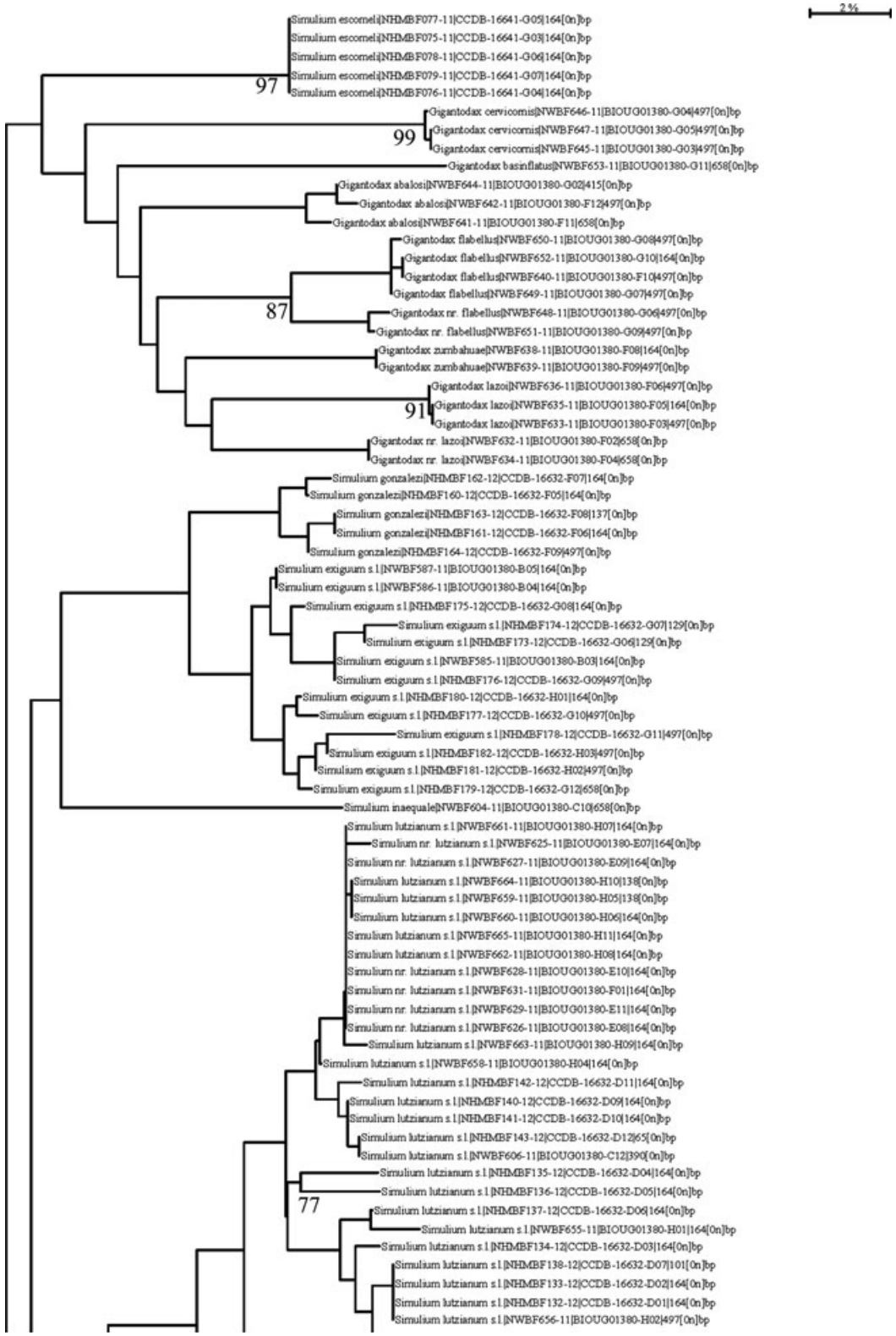

Fig. 2 Bootstrapped neighbour-joining (NJ) analysis tree composed of full-length (658 bp), ~300- to 400-bp and ~68- to 164-bp *Simuliidae* barcodes. A divergence of >2% is indicative of separate operational taxonomic units. Each specimen is labelled with species name (based on morphological ID) and corresponding barcode sequence length (in nucleotides). The number of ambiguous bases (N) is shown beside the sequence length in square parentheses. Bootstrapped values higher than 80% are shown in the tree below each node.

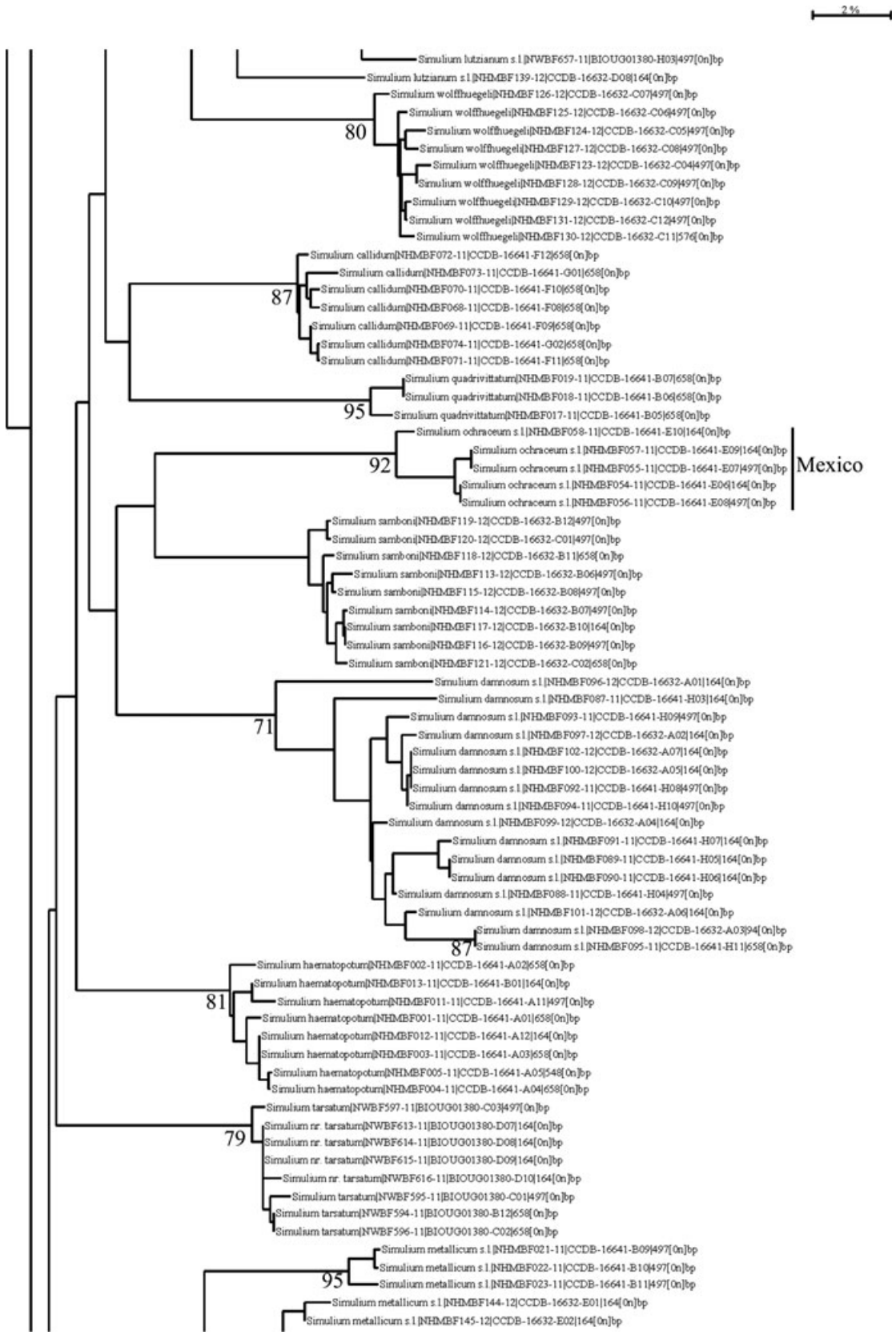

Fig. 2 (Continued)

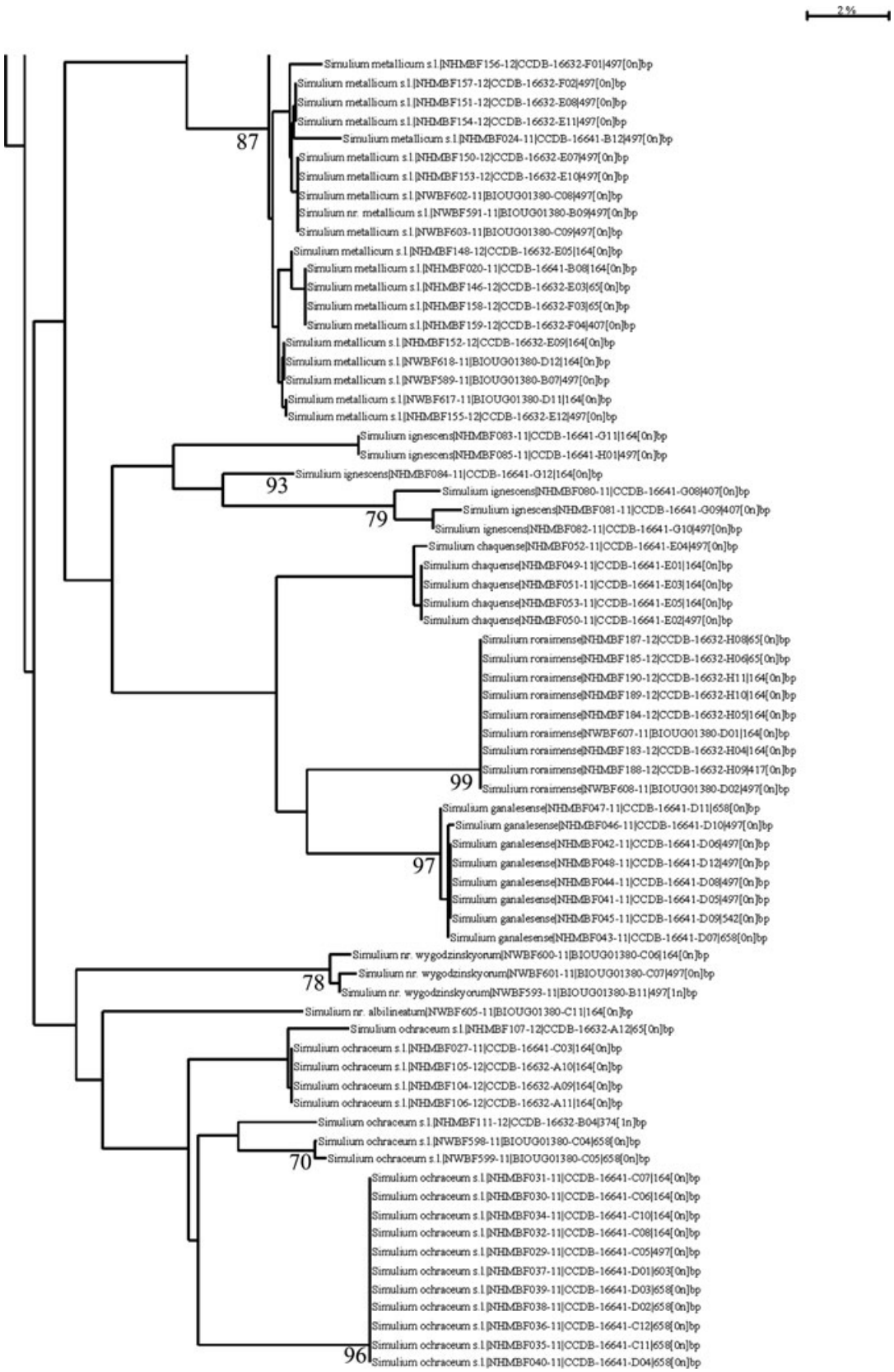

Fig. 2 (Continued)

multiple primer sets that generated various amplicon sizes. We started by targeting fragments of ~300–400 bp in size, which had the advantage of overlapping to form a full-length 658-bp COI barcode. Contrary to the general rule that the smaller the amplicon, the higher the success rate, the 388-bp primers produced approximately 22 more barcode records than the 307-bp primers, highlighting the importance of primer-template matching when working with degraded DNA. To recover sequence data from specimens that failed to amplify at least 300 bp, we also targeted fragments of 164 and 94 bp. Contrary to expectations, the 164-bp primers outperformed the 94-bp primers by more than 50%, but most of the 94-bp failures can be attributed to the fact that the 94-bp primers are prone to co-amplifying trace amounts of contaminating DNA. This is due to the fact that even small amounts of degraded DNA (i.e. contaminants) can easily persist as small (~100 bp) fragments for long period of time (Allentoft *et al.* 2012) and are thus more readily amplified by primer pairs that target within this size range. Additionally, the 94-bp amplicons are at the threshold of what can be cleanly sequenced on an ABI 3730XL capillary sequencer using standard BigDye sequencing chemistry (Applied Biosystems). Consequently, regions of low sequence quality had a high tendency to render traces unusable. Finally, the 94-bp primers may not bind to the simuliid target sequence as well as the 164-bp primers. The most probable explanation is that multiple factors acted in concert to inhibit recovery of high quality 94-bp fragments.

Regardless of how our blackfly DNA sequences were generated or how long the resulting sequences were, most sequences produced the expected patterns of clustering when combined in NJ analysis. That is, barcodes originating from the same species clustered tightly regardless of sequence length (Fig. 2). The only specimens that gave unexpected results were five specimens of *S. ochraceum s.l.*, all from Mexico (Fig. 2). Two of these five specimens had ~500-bp barcodes, whereas three had 164-bp barcodes. Examination of the barcode trace files and comparison of the barcode sequences to those from freshly collected *S. ochraceum s.l.* specimens (not shown) precluded the possibility of contamination and/or misidentification. The unique clustering of this subgroup of *S. ochraceum s.l.* specimens can be explained by their geographical provenance, as they might represent a different species within the *S. ochraceum* complex (see Adler & Crosskey 2013; Shelley *et al.* 2002). Further studies are underway to determine the full extent of genetic divergence within *S. ochraceum s.l.* populations.

This study has demonstrated that with the correct primers, these and probably other museum specimens of simuliids can be barcoded in a high-throughput manner. We have also shown that NJ analysis of short barcodes can correctly delineate simuliid species (with the exception of unresolved species complexes). Furthermore, this study highlights the importance of targeting the largest plausible amplicon to avoid co-amplification of nontarget DNA. If experimental parameters are optimized and the barcode target regions are carefully chosen, much useful data can be extracted from the rarely tapped biodiversity richness found in museum collections.

Acknowledgements

The authors thank the Diptera Curators at the Natural History Museum (NHM), London, UK, especially Zoe Adams, Erica McAlister and Theresa Howard, for their help in providing space and facilities in support of this study. This research was enabled by grants from Genome Canada through the Ontario Genomics Institute (OGI), the Canada Foundation for Innovation (CFI), the Ontario Ministry of Research and Innovation (MRI), the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Gordon and Betty Moore Foundation. Financial support was also provided by CONACYT–México (Grant Nos. 122424 and 194045) to Mario A. Rodríguez-Pérez.

References

Adler PH, Crosskey RW (2013) *World blackflies* (*Diptera: Simuliidae*): a comprehensive revision of the taxonomic and geographical inventory. Available from http://entweb. clemson.edu/biomia/pdfs/blackflyinventory.pdf. 120 pp. Accessed 11 July 2013.

Adler PH, Currie DC, Wood DM (2004) *The blackflies (Simuliidae) of North America*. pp. xv + 941 pp. Cornell University Press, Ithaca, New York.

Allentoft ME, Collins M, Harker D *et al.* (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 4724–4733.

Baird DJ, Pascoe TJ, Zhou X, Hajibabaei M (2011) Building freshwater macroinvertebrates DNA-barcode libraries from reference collection material: formalin preservation vs specimen age. *Journal of the North American Benthological Society*, **30**, 125–130.

Bluemel JK, King RA, Virant-Doberlet M, Symondson WOC (2011) Primers for identification of type and other archived specimens of *Aphrodes* leafhoppers (Hemiptera, Cicadellidae). *Molecular Ecology Resources*, **11**, 770–774.

Bucklin A, Hopcroft RR, Kosobokova KN *et al.* (2010) DNA barcoding of Arctic Ocean holozooplankton for species identification and recognition. *Deep Sea Research Part I: Topical Studies in Oceanography*, **57**, 40–48.

Conflitti IM, Kratochvil MJ, Spironello M, Shields GF, Currie DC (2010) Good species behaving badly: non-monophyly of black fly sibling species in the *Simulium arcticum* complex (Diptera: Simuliidae). *Molecular Phylogenetics and Evolution*, **57**, 245–257.

Currie DC, Adler PH (2008) Global diversity of blackflies (Diptera: Simuliidae) in freshwater. *Hydrobiologia*, **595**, 469–475.

Dean MD, Ballard JWO (2001) Factors affecting mitochondrial DNA quality from museum preserved *Drosophila simulans*. *Entomologia Experimentalis et Applicata*, **98**, 279–283.

Eaton DP, Diaz LA, Hans-Filho G *et al.* (1998) The Cooperative Group on Fogo Selvagem Research. 1998. Comparison of Blackflies species (Diptera: Simuliidae) on an Amerindian reservation with a high prevalence of fogo selvagem to neighbouring disease free sites in the state of Mato Grosso do Sul, Brazil. *Journal of Medical Entomology*, **35**, 120–131.

Edgar RC (2004) MUSCLE: multiple sequences alignment with high accuracy and high throughput. *Nucleic Acids Research*, **5**, 1792–1797.

Feld CK, Kiel E, Lautenschläger M (2002) The indication of morphological degradation of streams and rivers using Simuliidae. *Limnologica*, **32**, 273–288.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Hajibabaei M, McKenna C (2012) DNA Mini-barcodes. *Methods in Molecular Biology*, **859**, 339–353.

Hajibabaei M, Smith MA, Janzen DH, Rodríguez JJ, Whitfield JB, Hebert PDN (2006) A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, **6**, 959–964.

Hausmann A, Hebert PDN, Mitchell A *et al.* (2009) Revision of the Australian *Oenochroma vinaria* Guenée, 1858 species-complex (Lepidoptera: Geometridae, Oenochrominae): DNA barcoding reveals cryptic diversity and assesses status of type specimen without dissection. *Zootaxa*, **2239**, 1–21.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–321.

Hebert PDN, Ratnasingham S, DeWaard JR (2003b) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, **270**, S96–S99.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the Neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Science*, **12**, 14812–14817.

Hebert PDN, DeWaard JR, Zakharov EV *et al.* (2013) A DNA 'barcode blitz': rapid digitization and sequencing of a natural history collection. *PLoS ONE*, **8**, e68535.

Hernández LM (2007) Recommendations by The Natural History Museum on methods for collecting, preserving, rearing and mailing of simuliid specimens. Available from http://www.blackflies.info. Accessed 12 July 2012.

Hernández LM (2011) *Systematics of the blackfly subgenus Trichodagmia Enderlein* (*Diptera: Simuliidae: Simulium*) *in the New World*. PhD thesis, Wageningen University, Wöhrmann Printing Services, 546 pp.

Hernández-Triana LM, Crainey JL, Hall A *et al.* (2012) The utility of DNA barcoding for species identification within the blackfly Subgenus *Trichodagmia* Enderlein (Diptera: Simuliidae: *Simulium*) and related taxa in the New World. *Zootaxa*, **3514**, 43–69.

Ivanova NV, DeWaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, **6**, 998–1002.

Knölke S, Erlacher S, Hausmann A, Miller MA, Segerer AH (2005) A procedure for combined genitalia and DNA extraction in Lepidoptera. *Insect Systematics and Evolution*, **35**, 401–409.

Koch DA, Duncan GA, Parson TJ, Pruess KP, Powers TO (1998) Effects of preservation methods, parasites, and gut contents of blackflies (Diptera: Simuliidae) on Polymerase Chain Reaction products. *Journal of Medical Entomology*, **35**, 313–318.

Kumar NP, Rajavel AR, Natarajan R, Jambulingam P (2007) DNA barcodes can distinguish species of Indian mosquitoes (Diptera: Culicidae). *Journal of Medical Entomology*, **44**, 1–7.

Kumar NP, Srinivasan R, Jambulingam P (2012) DNA barcoding for identification of sand-flies (Diptera: Psychodidae) in India. *Molecular Ecology Resources*, **12**, 414–420.

Lautenschalger M, Kiel E (2005) Assessing morphological degradation in running water using Blackfly communities (Diptera, Simuliidae): can habitat quality be predicted from land use? *Limnologica*, **35**, 262–273.

Malmqvist B, Wotton RS, Zhang Y (2001) Suspension feeders transform massive amounts of seston in large northern rivers. *Oikos*, **92**, 35–43.

Malmqvist B, Adler PH, Kuusela K, Merrit RW, Wotton RS (2004) Blackflies in the boreal biome, key organisms in both terrestrial and aquatic environments; a review. *Ecoscience*, **11**, 187–200.

Mandriolli M, Borsatti F, Mola L (2006) Factors affecting DNA preservation from museum-collected lepidopteran specimens. *Entomologia Experimentali et Applicata*, **120**, 239–244.

Messing J (1993) M13 cloning vehicles. Their contribution to DNA sequencing. *Methods in molecular biology*, **23**, 9–22.

Meusnier I, Singer GAC, Landry J-F, Hickey DA, Hebert PND, Hajibabaei M (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, **9**, 214.

Pinheiro FP (1983). *Síndrome Hemorrágica de Altamira*, 2nd edn, pp. 107–108, Saúde na Amazônia, Brazil.

Pinheiro FP, Bensabath G, Costa D, Maroja OM, Lins Z, Andrade AMP (1974) Hemorrhagic syndrome of Altamira. *The Lancet*, **1**, 639–642.

Pinheiro FP, Bensabath G, Freitas RB, Costa D (1986) Fundação Serviços de Saúde Pública. Instituto Evandro Chagas: 50 anos de contribuição às ciências biológicas e à medicina tropical. *Fundação Serviços de Saúde Pública*, **2**, 795–798.

Pramual P, Kuvangkadilok C (2009) Agricultural land use and blackflies (Diptera, Simuliidae) species richness and species assemblages in tropical streams, Northern Thailand. *Hidrobiologia*, **625**, 173–184.

Pramual P, Wongpakam K, Adler PH (2011) Cryptic biodiversity and phylogenetic relationships revealed by DNA barcoding of Oriental black flies in the subgenus *Gomphostilbia* (Diptera: Simuliidae). *Genome*, **54**, 1–9.

Quicke DLJ, Belshaw R, Lopez-Vaamonde C (1999) Preservation of Hymenoptera specimens for subsequent molecular and morphological study. *Zoologica Scripta*, **28**, 261–267.

Ratnasingham S, Hebert PND (2007) BOLD: the barcode of life data system. (http://www. barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Rivera J, Currie D (2009) Identification of Nearctic blackflies using DNA barcodes (Diptera: Simuliidae). *Molecular Ecology Resources*, **9**, 224–236.

Saitou N, Nei M (1987) The Neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Schander C, Halanychi K (2003) DNA, PCR and formalinized animal tissue – a short review and protocols. *Organisms Diversity and Evolution*, **3**, 195–205.

Shelley AJ, Hernández LM, Penn M (2002) A biosystematic revision of the blackflies (Diptera: Simuliidae) of Belize, Central America. *Bulletin of the Natural History Museum Entomology*, **71**, 135–271.

Shelley AJ, Hernández LM, Maia-Herzog M, Luna DA, Garritano PR (2010). The Blackflies of Brazil (Diptera, Simuliidae). In: *Aquatic Biodiversity in Latin America (ABLA Series)*, **Vol. 6**. (eds Adis J, JArias J, S. Golovatch S, Mantzev KM, Rueda-Delgado G & Domínguez E), pp. **814**. Pensoft, Sofia-Moscosw.

Shokralla S, Zhou X, Janzen DH *et al.* (2011) Pyrosequencing for mini-barcoding of fresh and old museum specimens. *PLoS ONE*, **7**, 1–4.

Tamura K, Peterson D, Peterson N, Stoecher G, Nei M, Kumar S (2007) MEGA 5. Available from http://www.megasoftware.net. Accessed 15 October 2013.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **4**, 1–18.

Van Houdt JKJ, Breman FC, Virgilio M, De Meyer M (2010) Recovering full DNA barcodes from natural history collections of Tephritid fruitflies (Tephritidae, Diptera) using mini barcodes. *Molecular Ecology Resources*, **10**, 459–465.

Watts PC, Thompson DJ, Allen KA, Kemp SJ (2007) How useful is DNA extracted from the legs of archived insects for microsatellite-based population genetic analyses? *Journal of Insect Conservation*, **11**, 195–198.

Zimmerman J, Hajibabaei M, Blackburn DC *et al.* (2008) DNA damage in preserved specimens and tissue samples: a molecular assessment. *Frontiers in Zoology*, **5**, 1–13.

L.M.H.T., S.P., M.A.R.P. and L.G.C. designed the study; L.M.H.T. and S.P. cowrote the first version of the manuscript; S.W.P. carried out the DNA sequencing and together with L.M.H.T. performed sequencing analysis; P.D.N.H. contributed logistical support for the research reported here and contributed to data interpretation and co-authored the manuscript and T.R.G. contributed to data interpretation and co-authored the manuscript.

Data Accessibility

DNA sequences: GenBank Accessions nos. KF839950 – KF840065.

BOLD project title and project code: 'Old samples of blackflies (Simuliidae) held at the NHM Collection_2011_2012'; code: NHMBF.

DOI link in BOLD: dx.doi.org/10.5883/DS-NHMBF13.

Sample information and sequence alignments: Supplemental Material for Online Publication.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1 Old samples of blackflies (Simuliidae) held at the NHM collection_2011, 2012: Project Code, Sample and Process ID, GenBank Accession Number, and Species Identification.

Data S2 DNA barcoding sequences alignment for old samples of blackflies held at the NHM collection 2011_2012.